



OPEN ACCESS

EDITED BY

Tse-Yen Yang,
Taipei Medical University, Taiwan

REVIEWED BY

Rocio Gonzalez-Soltero,
European University of Madrid, Spain
TaChen Chen,
Nihon Pharmaceutical University, Japan
Clair Blacketer,
Janssen Research and Development,
United States

*CORRESPONDENCE

Brenda Mbouamba Yankam
✉ brenda.yankammbouamba@ruhr-uni-bochum.de;
✉ brenda.yankam@gmail.com

Evaluating the impact of OMOP-CDM on data quality insight generation in respiratory disease management

Brenda Mbouamba Yankam^{1,2*},
Fankoua Tchaptchet Luc Baudoin¹, Pauline Andeso³,
François Anicet Onana Akoa^{1,4}, Jean Blaise Ebimbe⁵,
Miranda Barasa³, Mbele Onana⁶, Samuel Iddi^{3,7}, Agnes Kiragga^{3,8},
Bertrand Hugo Mbatchou Ngahane^{1,9} and Data Science Without
Borders Project

¹Centre for Health Data Science Without Borders Project, Centre for Health Data Science Without Borders, Cameroon; ²Faculty of



Evaluating the Impact of OMOP-CDM on Data Quality Insight Generation in Respiratory Disease Management

**Presented by:
Dr Brenda Yankam**



Presentation Outline

- Introduction
- Solution
- Objective
- Methodology
- Data quality assessment with DQD and Achilles
- Result
- Challenges encountered



Introduction

- **Why does this matter?**
- Healthcare data are increasingly available from:
 - Electronic Health Records; Registries; Hospital databases
- Data are often:
 - Heterogeneous; Unstructured; Stored in different formats
- Poor interoperability limits reuse for research and decision-making



Solution

- **OMOP Common Data Model (CDM)**
 - Standardizes healthcare data
 - Enables multi-center research
 - Supports reproducible analytics
- ❖ How do we ensure that standardized data are actually of high quality?



Objective

- To evaluate the impact of OMOP-CDM standardization on generating meaningful data quality insights from a respiratory disease dataset.

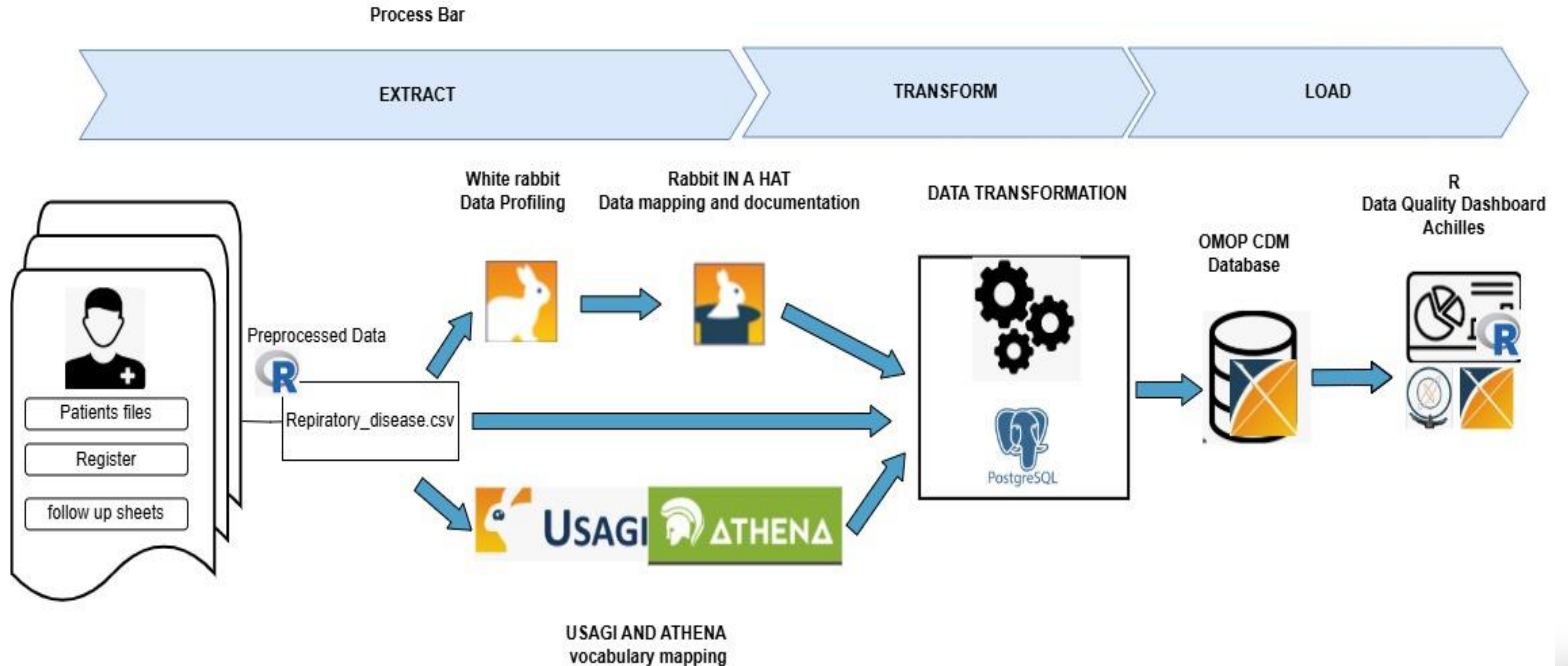


Methodology

Dataset Description

- In paper-based form, and in French was digitized, and anonymized.
- The data from 2009 to 2023, comprised 108 variables and 2,154 patient records.
- The variables in this dataset was categorised as follows:
 - Demographics; clinical history; symptoms; laboratory, measurements and outcomes

OMOP CDM Standardization Workflow



Vocabulary Mapping with Usagi

Usagi - Res_usagi_Final.csv

File Edit View Help

Status	Source c...	Source term	Frequency	Match sc...	Concept ...	Concept ...	Domain	Concept ...	Vocabulary	Concept ...	Standard...	Parents	Children	Assigne...	Equival...	Co...	Stat...
Approved	13	Cavum		0.64	4141446	Cavum v...	Spec An...	Body Str...	SNOMED	33191008	S	2	1				
Approved	5	CAP		0.53	763657	Caper	Observat...	Substance	SNOMED	4444610...	S	1	0				
Approved	4	Cancer		1.00	40757663	Cancer	Observat...	Survey	LOINC	54532-7	S	3	1				
Approved	2	C4		1.00	36307912	Hospital ...	Meas Val...	Answer	LOINC	LA26181-0	S	0	0				
Approved	1	C3		1.00	36307912	Hospital ...	Meas Val...	Answer	LOINC	LA26181-0	S	0	0				
Approved	11	Buttock sores		0.78	37159955	Ulcer of...	Condition	Disorder	SNOMED	1067219...	S	2	4				
Approved	11	Bronchitis		1.00	256451	Bronchitis	Condition	Disorder	SNOMED	32398004	S	3	19				
Approved	15	Bronchiectasis		1.00	256449	Bronchie...	Condition	Disorder	SNOMED	12295008	S	1	16				
Approved	25	Bronchiectasis		1.00	256449	Bronchie...	Condition	Disorder	SNOMED	12295008	S	1	16				
Approved	8	Bronchial		0.87	4111805	Squamo...	Condition	Disorder	SNOMED	2546220...	S	4	6				
Approved	10	Breast cancer		1.00	4112853	Malignan...	Condition	Disorder	SNOMED	2548370...	S	2	23				
Approved	4	Breast		1.00	4298444	Breast st...	Spec An...	Body Str...	SNOMED	76752008	S	3	7				
Approved	2	Breast		1.00	4298444	Breast st...	Spec An...	Body Str...	SNOMED	76752008	S	3	7				
Approved	7	Bone marrow		1.00	4029619	Bone ma...	Spec An...	Body Str...	SNOMED	14016003	S	4	5				
Approved	10	AVC and complication		0.50	4293463	Commu...	Condition	Disorder	SNOMED	3850930...	S	1	0				
Approved	6	Asthma exacerbation		0.87	1340255	Exacerba...	Condition	Disorder	OMOP Ex...	OMOP51...	S	2	0				
Approved	8	Asthma		1.00	317009	Asthma	Condition	Disorder	SNOMED	1959670...	S	1	31				

Source code

Source code	Source term	Frequency
8	Asthma	-1

Target concepts

Concept ID	Concept name	Domain	Concept class	Vocabulary	Concept code	Standard concept	Parents	Children	Mapping Type	Creation Proven...
317009	Asthma	Condition	Disorder	SNOMED	195967001	S	1	31	MAPS_TO	

MAPS_TO Remove concept

Search

Query

Use source term

Query:

Filters

Filter by user selected concepts / ATC code

Filter standard concepts

Include source terms

Filter by concept class:

Filter by vocabulary:

Filter by domain:

Results

Score	Term	Concept ID	Concept name	Domain	Concept class	Vocabulary	Concept code	Standard concept	Parents	Children
0.77	Seasonal asthma	40483397	Seasonal asthma	Condition	Disorder	SNOMED	445427006	S	2	1
0.77	Mild asthma	4155468	Mild asthma	Condition	Disorder	SNOMED	370218001	S	1	2
0.76	Asthma finding	4308356	Asthma finding	Condition	Clinical Finding	SNOMED	390798007	S	1	23
0.76	Moderate asthma	4155469	Moderate asthma	Condition	Disorder	SNOMED	370219009	S	1	1
0.75	Severe asthma	4152913	Severe asthma	Condition	Disorder	SNOMED	370221004	S	1	3

MAPS_TO Replace concept Add concept

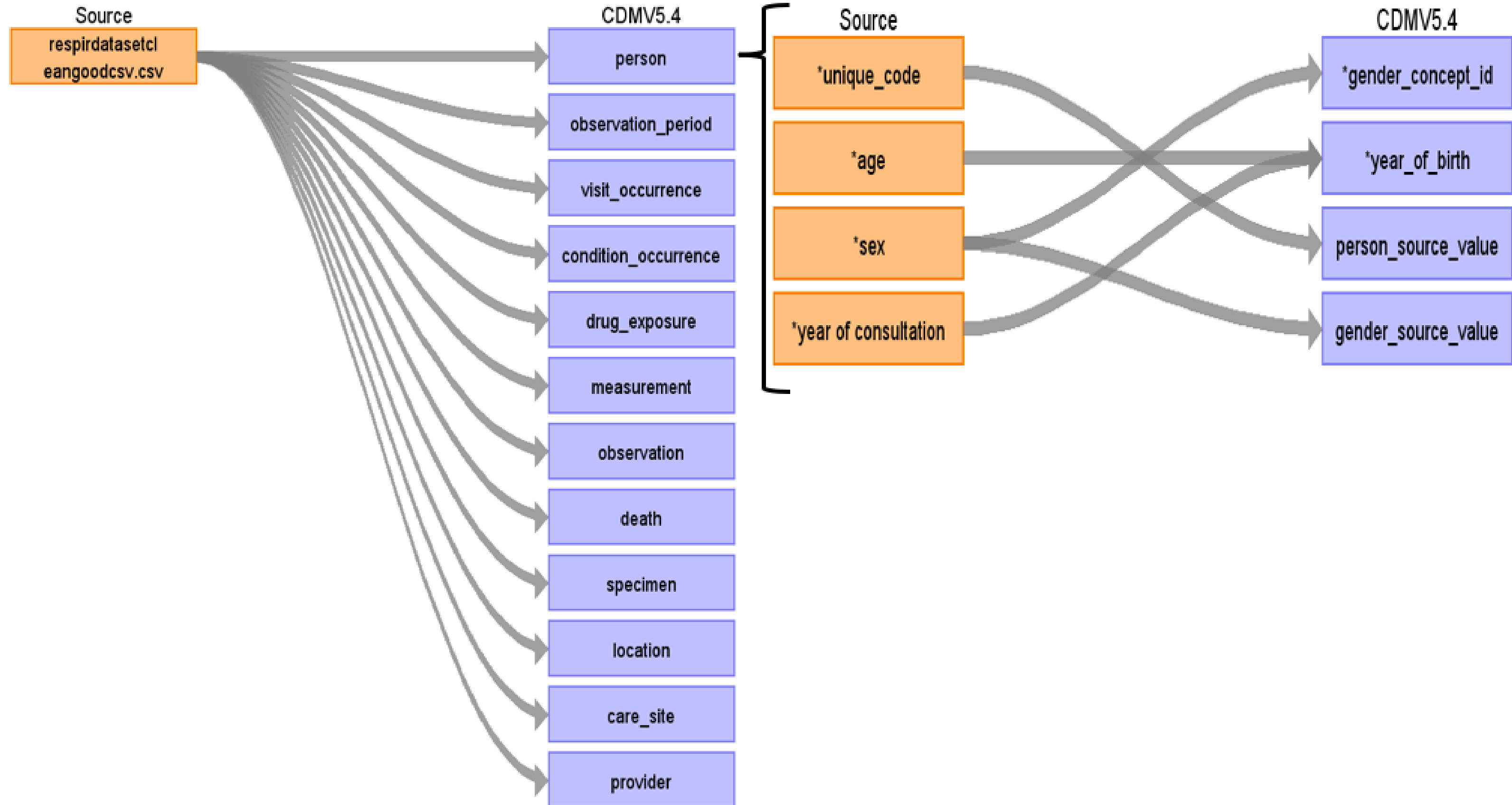
Comment:

Flag EQUAL Unapprove

Approved / total: 201 / 270 74.4% of total frequency

Author: Yankam Brenda Vocabulary version: v5.0 30-AUG-24

ETL Mapping of Source Data to OMOP CDM Person Table in Rabbit in a Hat





Data Quality Assessment

- DQD and Achilles.
- DQD validation dimensions:
 - Conformance: Does data follow OMOP standards?
 - Completeness: Are values present?
 - Plausibility: clinical validity of values

Data Quality Assessment Before ETL Verification of Failed Checks



RESPIRATORY DISEASES INPATIENTS

OVERVIEW

METADATA

RESULTS

ABOUT

DATA QUALITY ASSESSMENT

RESPIRATORY DISEASES INPATIENTS

DataQualityDashboard Version: 2.6.1

Results generated at 2025-05-15 13:26:06 in 1 mins

	Verification				Validation				Total			
	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
Plausibility	497	16	513	97%	291	0	291	100%	788	16	804	98%
Conformance	880	46	926	95%	137	8	145	94%	1017	54	1071	95%
Completeness	448	4	452	99%	16	1	17	94%	464	5	469	99%
Total	1825	66	1891	97%	444	9	453	98%	2269	75	2344	97%

1439 out of 2269 passed checks are Not Applicable, due to empty tables or fields.

39 out of 75 failed checks are SQL errors.

Corrected pass percentage for NA and Errors: 96% (830/866).

Data Quality Assessment After ETL Verification of Failed Checks



DOUALA GENERAL HOSPITAL CDM

OVERVIEW

METADATA

RESULTS

ABOUT

DATA QUALITY ASSESSMENT

DOUALA GENERAL HOSPITAL CDM

DataQualityDashboard Version: 2.6.1

Results generated at 2026-01-08 10:32:53 in 2 mins

	Verification				Validation				Total			
	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
Plausibility	503	10	513	98%	291	0	291	100%	794	10	804	99%
Conformance	899	27	926	97%	137	8	145	94%	1036	35	1071	97%
Completeness	448	4	452	99%	16	1	17	94%	464	5	469	99%
Total	1850	41	1891	98%	444	9	453	98%	2294	50	2344	98%

1471 out of 2294 passed checks are Not Applicable, due to empty tables or fields.

39 out of 50 failed checks are SQL errors.

Corrected pass percentage for NA and Errors: 99% (823/834).

Challenges Encountered

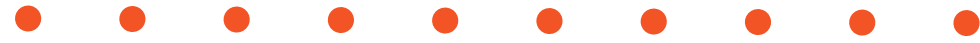
Challenge	Description	Solution
Mapping Past Medical Histories	Difficulty in mapping past medical histories as conditions, as OMOP is designed for current conditions .	Used a non-standard Condition Type Concept ID to differentiate past conditions from present ones.
Absence of ID Codes for Local Professions	Local professions in Cameroon (e.g., specific nurses or technicians) lacked predefined codes in OMOP's standard vocabulary.	No solution was found.
Measurement Units	Several medical measurements (e.g., creatinine, respiratory rate) lacked standard units, creating inconsistencies.	Consulted with a domain expert to standardize the measurement units according to medical conventions.

Partners



African Population and Health Research Center
Transforming lives in Africa through research.





Thank You



**Data Science
Without Borders**

